# The Development and Determination of Chemically Distinct Solute Parameters for Use in Linear Solvation Energy Relationships

Jeff D. Weckwerth
Hutchinson Technology Incorporated, 40 W. Highland Park,
Hutchinson, Minnesota 55350

Mark F. Vitha
Drake University, Department of Chemistry, 2507 University Avenue,
Des Moines, Iowa 50311

Peter W. Carr*
University of Minnesota, Department of Chemistry, Smith and Kolthoff Halls,
207 Pleasant St. S.E., Minneapolis, Minnesota 55455

* Author to whom correspondence should be addressed.

**Abstract**

Linear solvation energy relationships (LSERs) have been used to correlate hundreds of chemical systems. However, several of the solute parameters used in the LSERs do not model any one specific type of interaction, but rather are blends of multiple interactions. Thus, LSERs generated with these parameters cannot be rigorously interpreted to provide a complete understanding of the forces underlying solute-solvent interactions. In this work, the development of a new set of chemically distinct solute parameters based on Kovats retention indices on a variety of GC stationary phases is presented. The determination of the parameters does not utilize the same initial input estimates as to the values of the parameters which were used in previous studies and which were based on descriptors that reflected a blend of molecular properties. Using this new method, a set of parameters describing the dipolarity, polarizability, size, and hydrogen bond acidity of 53 compounds has been determined. The parameters are shown to be superior to older parameters in that they are chemically "purer" (i.e. are not blends of multiple interactions). The success of these new parameters in LSERs for the correlation of gas/water and water/octanol partitioning, and retention in gas chromatography is also presented.

## Introduction

The solvatochromic comparison method, as originally developed by Kamlet, Taft, et al., has been used quite successfully to correlate hundreds of chemical systems [1]. The recent improvements to the solute scales by Abraham et al. and the measurement of these solute parameters for over 1000 compounds allow for excellent correlations and accurate predictions for many important chemical systems [2-4]. Additionally, as with all linear solvation energy relationships (LSERs), the signs and magnitudes of the coefficients provide important chemical information about the systems [5].

However, there are significant, fundamental problems with the interpretation of the LSERs. The fact that we need two different equations to model two different types of solution processes is symptomatic of these problems. The two equations (using Abraham's parameters) are as follows [6]:

$$SP = SP_0 + l\log L^{16} + s\,{}_2^H + a\,{}_2^H + b\,{}_2^H + rR_2, \qquad (1)$$

$$SP = SP_0 + vV_X + s\,{}_2^H + a\,{}_2^H + b\,{}_2^H + rR_2. \qquad (2)$$

Equation 1 is used for various gas-to-liquid transfers, such as retention in gas chromatography (GC) [6], while equation 2 is used for processes between condensed phases [7]. Generally, whenever water is one of the solvents, equation 2 must be used [8].

While the hydrogen bonding terms ( ${}_2^H$ and ${}_2^H$) are very well defined and successfully model the respective interactions, the remaining terms ($V_X$, $\log L^{16}$, ${}_2^H$, and $R_2$) do not model any one *single* specific type of interaction. Both $V_X$ and $\log L^{16}$ are used to describe a combination of dispersion interactions and the cavity formation process [9]. As

yet, we cannot quantify how much of a *l* or *v* coefficient is due to dispersion and how much is due to cavity formation. Abraham, et al. used scaled-particle theory (SPT) in an attempt to answer these questions, but SPT cannot be easily applied to all systems [9]. Although $V_X$ and log $L^{16}$ are fairly closely related, substituting one for the other results in changes not only in the *v* and *l* coefficients as expected, but also in the *s* and *r* coefficients [4, 6]. This causes difficulty in the comparison of two systems when different equations are used.

Dispersion interactions are strongly dependent on solute polarizability, and cavity formation is dependent on solute volume [10]. Although polarizability and volume are very closely related, it is impossible to accurately model the two properties with only one parameter. Therefore the "polarizability correction factor" $R_2$ must be included in the LSER [10]. This is an excess polarizability term, defined as the difference between the polarizability of the compound and the polarizability of a hypothetical alkane of the same volume [10]. The purpose and the importance of this descriptor was analyzed in detail in earlier work [11]. Interpretation of the *r* coefficient is difficult, and in many cases the coefficient is simply ignored, with no attempt made at interpretation. The sign of the coefficient can be either positive or negative, and the magnitude is usually small, but statistically significant [6].

The situation is further complicated by the fact that the $\pi_2^H$ term includes a significant contribution from solute polarizability in addition to dipolarity [12, 13]. Both the Abraham and the Carr groups have been unsuccessful in their attempts to remove the polarizability component from their respective solute dipolarity terms, $\pi_2^H$ and $\pi{*_2}^C$ [12,

3

13]. Finally, bringing this convolution of interactions full circle, the log $L^{16}$ term must include a small dipolarity contribution resulting from induction interactions between dipolar solutes and hexadecane. Meyer has shown that induction interactions are much more important than commonly believed [14].

Theoretical Basis for Parameter Development

In this work we create a new set of solute parameters which are more chemically specific than the current sets of parameters. In particular, we want to define a pure dipolarity term that has no polarizability contribution. The polarizability inherent in both the $_2^H$ and the $*_2^C$ may result from the manner in which these scales were developed. Initial estimates for each of these descriptors were taken from the Kamlet-Taft solvent $*$ scale, which is known to include a significant polarizability contribution [15, 16]. In our development of a solute dipolarity scale ($D_X$), we will make no initial estimates of solute dipolarity. Instead we will use the equation for GC retention,

$$log\ k_X\ =\ log\ \ +\ v\ V_X\ +\ p\ P_X\ +\ d\ D_X\ +\ a\ A_X\ +\ b\ B_X, \tag{3}$$

and the related equation for the Kovats retention index ($I_X$) of a solute,

$$I_X\ =\ -\frac{z_1}{z_2}\ +\frac{1}{z_2}\ V_X\ +\ \frac{p}{v\ z_2\ +\ p\ z_4}\ R_X\ +\ \frac{d}{v\ z_2\ +\ p\ z_4}\ D_X\ +\ \frac{a}{v\ z_2\ +\ p\ z_4}\ A_X\ +\ \frac{b}{v\ z_2\ +\ p\ z_4}\ B_X, \tag{4}$$

both of which were derived in previous work [11]. In equation 4, $V_X$ is the solute volume (a cavity formation term), $D_X$ is the solute dipolarity, $A_X$ is the solute acidity, $B_X$ is the solute

basicity, and $R_X$ is the solute excess polarizability, which is related to the polarizability ($P_X$) through the following equation:

$$R_X = P_X - z_3 - z_4 \frac{V_X - z_1}{z_2} \tag{5}$$

The constants, $z_1$, $z_2$, $z_3$, and $z_4$ result from the relationships between $V_X$, $P_X$, and $N$ (number of carbon atoms) for $n$-alkanes, as given by

$$V_{n\text{-alkane}} = z_1 + z_2 N \quad \text{and} \quad P_{n\text{-alkane}} = z_3 + z_4 N \tag{6}$$

## Experimental

The details of the GC retention database used in this study have been published elsewhere [17]. The data set included logarithms of retention factors (log $k$) for 53 compounds on 8 stationary phases at several different temperatures. One of the stationary phases, polyethylene glycol, was excluded from the data set because several data points were missing. Rather than including the entire 29 column data matrix in the regressions, we have used only the data at 80 °C for each stationary phase. Outliers were detected with the least median of squares regression program, PROGRESS [18]. All other calculations and regressions were done with MATLAB [19].

## Results and Discussion

### Derivation of the New Parameters

We make the following definitions to simplify equation 4:

$$p' = \frac{p}{v \, z_2 + p \, z_4}, \quad d' = \frac{d}{v \, z_2 + p \, z_4}, \quad \text{and} \quad a' = \frac{a}{v \, z_2 + p \, z_4} \tag{8}$$

Because the stationary phases studied here have no hydrogen bond acidity, $b$ is equal to zero. With these simplifications, equation 4 becomes

$$I_X = -\frac{z_1}{z_2} + \frac{1}{z_2} V_X + p' \, R_X + d' \, D_X + a' \, A_X. \tag{8}$$

Rather than making initial estimates for $V_X$, $R_X$, $D_X$, and $A_X$, we use three reference compounds which are chemically well-understood. Cyclohexane is a nonpolar, non-hydrogen bond acidic reference compound, (i.e. $D_X = A_X = 0$ in equation 8), benzonitrile is a dipolar, but non-hydrogen bond acidic reference compound (i.e. $A_X = 0$) and phenol is a dipolar hydrogen bond acid (i.e. all parameters are non-zero). Dipolarities for all compounds will be calculated relative to the dipolarity of benzonitrile, and hydrogen bond acidities will be calculated relative to the acidity of phenol. Therefore, we define each of these values as 1, that is $D_{benzonitrile} = 1.0$ and $A_{phenol} = 1.0$.

By writing equation 8 for each reference compound, we can solve for each of the stationary phase-dependent coefficients, $p'$, $d'$, and $a'$, as follows:

$$p' = \frac{I_{cyclohexane} + \dfrac{z_1}{z_2} - \dfrac{1}{z_2} V_{cyclohexane}}{R_{cyclohexane}}, \tag{9}$$

$$d = I_{benzonitrile} + \frac{z_1}{z_2} - \frac{1}{z_2} V_{benzonitrile} - p \ R_{benzonitrile}, \tag{10}$$

$$a = I_{phenol} + \frac{z_1}{z_2} - \frac{1}{z_2} V_{phenol} - p \ R_{phenol} - d \ D_{phenol}. \tag{11}$$

Each of these coefficients can then be substituted back into equation 8 to give

$$I_X = C_1 + C_2 \ I_{cyclohexane} + C_3 \ I_{benzonitrile} + C_4 \ I_{phenol}, \tag{12}$$

where it can be shown that the variables $C_1$, $C_2$, $C_3$, and $C_4$ are related to the sought-after

solute parameters by the following equations:

$$C_4 = A_X. \tag{13}$$

$$C_3 = D_X - D_{phenol} \ C_4 \tag{14}$$

$$C_2 = \frac{R_X - R_{benzonitrile} \ C_3 - R_{phenol} \ C_4}{R_{cyclohexane}} \tag{15}$$

$$C_1 = \frac{V_X - z_1 - \left(V_{cyclohexane} - z_1\right) C_2 - \left(V_{benzonitrile} - z_1\right) C_3 - \left(V_{phenol} - z_1\right) C_4}{z_2} \tag{16}$$

It is important to note that the expressions for $C_1$, $C_2$, $C_3$, and $C_4$ include no properties

which are dependent on the stationary phase.  Therefore we can calculate these variables by

linear regression.  For example, we have the retention indices for methanol, cyclohexane,

benzonitrile, and phenol on several of the same stationary phases.  Regressing $I_{methanol}$

against $I_{cyclohexane}$, $I_{benzonitrile}$, and $I_{phenol}$ will give the intercept, $C_1$, and the coefficients, $C_2$,

$C_3$, and $C_4$.  These coefficients are specific to methanol, and they are related to the

molecular properties of methanol.  The same regression can be done for any compound of

interest to get a set of compound-specific variables, $C_1$, $C_2$, $C_3$, and $C_4$.

Calculation of $A_X$ and Comparison with Other Acidity Scales

Some molecular properties can be determined directly from these variables. For example, according to equation 17, the hydrogen bond acidity ($A_X$) of a compound is given by $C_4$. The results are given in Table 1. An acidity scale based on principle components analysis (PCA) was also developed in earlier work [11]. While the agreement between $A_X$ calculated by these two different methods is acceptable, the current method is more robust, and the values should be trusted more. Figure 1 compares the current $A_X$ with the hydrogen bond acidity parameters of Carr ($\alpha_2^{C}$) and Abraham ($\alpha_2^{H}$). Excluding acetic acid, agreement between $A_X$ and $\alpha_2^{C}$ is excellent. The agreement between $A_X$ and $\alpha_2^{H}$ is not as good. Presumably this arises from the fact that the same data set used to determine $A_X$ was used in part to determine $\alpha_2^{C}$ [12], whereas a different data set was used to determine $\alpha_2^{H}$.

In the determination of both $\alpha_2^{C}$ and $\alpha_2^{H}$, values of log $L^{16}$ for each compound were required [12, 13]. If there is any error in log $L^{16}$, this could show up as an error in $\alpha_2^{C}$ and $\alpha_2^{H}$. Acetic acid is especially problematic since carboxylic acids dimerize extensively in nonpolar solvents and even in the gas phase [12], and their retention may be influenced by gas-liquid interfacial adsorption or by adsorption on the diatomaceous earth support [12]. Symptomatic of these problems, Li et al. reported asymmetric peaks for carboxylic acids in the measurement of $L^{16}$, and the peak maximum shifted with the amount injected [12]. The current regression-based method for the determination of $A_X$ is unique in that it does not require values of log $L^{16}$, nor does it require initial estimates of $A_X$.

Calculation of $V_X$

Rearranging equation 20 we find the following equation for the volume of a compound:

$$V_X - z_2\,C_1 - z_1\left(1 - C_2 - C_3 - C_4\right) = V_{cyclohexane}\,C_2 + V_{benzonitrile}\,C_3 + V_{phenol}\,C_4. \quad (17)$$

We see no way to predict $V_X$ without making any further assumptions or initial estimates. We choose to use McGowan's characteristic volume ($V_M$) as an initial estimate for the volume scale [20]. McGowan's volume scale is based on estimates of the van der Waals volume and can be easily calculated from molecular structure [20]. To remove an intercept in subsequent correlations of $\mathrm{Log}L^{16}$ values, 6.56 was added to the McGowan volume for each compound. We refer to this adjusted volume as $V_M'$. We can calculate $z_1$ and $z_2$ by linear regression of $V_M'$ against the number of carbons for *n*-alkanes, as described by equation 6. Using equation 16, we can determine $V_{cyclohexane}$, $V_{benzonitrile}$, and $V_{phenol}$ by linear regression of the quantity $V_M' - z_2\,C_1 - z_1\,(1 - C_2 - C_3 - C_4)$ against $C_2$, $C_3$, and $C_4$ for all compounds. These values can then be substituted back into equation 17 to calculate our volume term, $V_X$. The values of $V_X$ are given in Table 1 and are compared to $V_M'$ in Figure 2.

Calculation of $P_X$

Combining equations 15, 5, and 17, we arrive at the following equation for the polarizability of a compound:

$$P_X - z_4\,C_1 - z_3\left(1 - C_2 - C_3 - C_4\right) = P_{cyclohexane}\,C_2 + P_{benzonitrile}\,C_3 + P_{phenol}\,C_4. \quad (18)$$

Like the volume term, we have been unable to predict $P_X$ without making any further assumptions or initial estimates. A common measure of polarizability is the molar refractivity (*MR*) [10],

$$MR = \frac{n^2 - 1}{n^2 + 2} \bar{V} \,, \tag{19}$$

which is related to the refractive index (*n*) and molar volume ($\bar{V}$) of a compound. However, this property must be determined experimentally. Even though values of *MR* are available for hundreds of compounds, we would like to use a parameter which can be easily calculated from the molecular structure of any compound, such as the polarizability scale of Miller and Savchik ($P_{M\text{-}S}$) [21], which is closely related to the molar refractivity [11].

Just as we did for the volume scale, we can calculate $z_3$ and $z_4$ by linear regression of PM-S against the number of carbons for *n*-alkanes (equation 6). Using equation 18 we can determine $P_{cyclohexane}$, $P_{benzonitrile}$, and $P_{phenol}$ by linear regression of the quantity, $P_{M\text{-}S} - z_4 \, C_1 - z_3 \, (1 - C_2 - C_3 - C_4)$, against $C_2$, $C_3$, and $C_4$ for all compounds. These values can then be substituted back into equation 18 to calculate our polarizability term, $P_X$. The values of and $P_X$ are given in Table 1 and compared to $P_{M\text{-}S}$ in Figure 3.

Calculation of $D_X$

From equation 14 we know that the dipolarity of a compound ($D_X$) is related to $C_3$ and $C_4$, but we do not yet know the value of $D_{phenol}$. We can, however, calculate $D_X$ for non-hydrogen bond acidic compounds, for which $C_4$ is equal to zero, such that $C_3 = D_X$.

The only task remaining is the calculation of $D_X$ for hydrogen bond donors. To accomplish this, we use the values of $\log L^{16}$ which have been carefully measured for each of the compounds in this study. The equation for a gas-to-liquid partition coefficient,

$$\log K_X = v V_X + p P_X + d D_X + a A_X + b B_X \qquad (20)$$

was derived in earlier work [11]. We know that hexadecane is non-acidic and non-basic, therefore $a$ and $b$ are equal to zero, and equation 20 becomes

$$\log L^{16} = v V_X + p P_X + d D_X. \qquad (21)$$

Substituting equation 14 in this expression gives:

$$\log L^{16} = v V_X + p P_X + d C_3 + \left(d D_{phenol}\right) C_4 = v V_X + p P_X + d C_3 + c C_4, \qquad (22)$$

where $c = d D_{phenol}$. The coefficients $d$ and $c$ are determined from a linear regression of $\log L^{16}$ against $V_X$, $P_X$, $C_3$, and $C_4$. The results of the regression described by equation 22 are:

$$\log L^{16} = -0.0456 V_X + 0.620 P_X + 0.994 C_3 + 0.602 C_4 \qquad (24)$$

$$R = 0.999 \qquad \overline{sd} = 0.052 \qquad N = 53.$$

According to this regression, d = 0.994 and c = 0.602. From these coefficients we can calculate that $D_{phenol} = 0.606$. This value, along with the values of $C_3$ and $C_4$ for each compound, are used in equation 14 to calculate $D_X$ for all compounds.

Finally, with all the parameters now defined, we re-scale the $D_X$ and $A_X$ values so that the extremely dipolar dimethylsulfoxide has a dipolarity of 1, and the highly acidic 1,1,1,3,3,3-hexafluoro-2-propanol has an acidity of 1.

11

## Analysis of log $L^{16}$ in Terms of Cavity Formation, Dispersion, and Induction

In many applications of LSERs, log $L^{16}$ is used as a descriptor to account for a combination of dispersion interactions and the cavity formation process. However, the relative importance of each of these processes is uncertain. In an attempt to explain the various contributions to log $L^{16}$, Abraham and Fuchs [2] presented the correlation:

$$log\ L^{16} = \left(0.057 \pm 0.095\right) - \left(4.56 \pm 0.74\right)\frac{V_I}{100} + \left(0.195 \pm 0.015\right)MR + \left(0.065 \pm 0.007\right)\mu^2 \qquad (24)$$

$$R = 0.973 \qquad \overline{sd} = 0.26 \qquad N = 84,$$

where $V_I$ is Leahy's intrinsic volume, $MR$ is the molar refractivity, and $\mu$ is the dipole moment. Chemically, the coefficients seem to make sense. The coefficient of the volume term is negative due to the endoergic cavity formation, while the coefficients of $MR$ and $\mu^2$ are positive, representing the exoergic dispersion and induction interactions, respectively. However, compared with many other LSERs in which log $L^{16}$ is an explanatory parameter, the quality of the fit is fairly poor. Dallas and Carr tried to improve the correlation by using various other descriptors [22]. Their most successful correlation was

$$log\ L^{16} = \left(-3.40 \pm 0.11\right) + \left(4.13 \pm 0.08\right)\frac{V_I}{100} + \left(18.55 \pm 0.71\right)\ \left(n\right) + \left(0.609 \pm 0.055\right)\ {}^*\mathcal{C}_2 \quad (25)$$

$$R = 0.994 \qquad \overline{sd} = 0.101 \qquad N = 96,$$

where $(n)$, the Onsager function, is defined by

$$\left(n\right) = \frac{n^2 - 1}{2\ n^2 + 1}. \qquad (26)$$

The use of $(n)$ and $*_2^C$ significantly improved the quality of the fit, making the equation more suitable for predictive purposes. However, these coefficients do not make much sense chemically [22].

With the regression-based descriptors which have been defined in this work, we are able to improve the quality of the correlation, without losing chemical information. The correlation is as follows:

$$log\ L^{16} = -0.0456\ V_X + 0.620\ P_X + 1.523\ D_X. \tag{27}$$

$$R = 0.999 \qquad \overline{sd} = 0.051 \qquad N = 53$$

Although $\log L^{16}$ was used to a small extent to define these parameters, this does not affect the quality of the regression. The only effect that this will have on the correlation is to force both the intercept and the $a$ coefficient to zero. The quality of the fit is excellent, especially compared to equation 24 and 25, but also by the LSER standards (see Figure 4).

We can use equation 27 to determine the absolute and relative importance of cavity formation, dispersion, and induction to $\log L^{16}$. This analysis has been presented elsewhere [11]. As expected, dispersive interactions dominate in $\log L^{16}$. While the contribution of inductive interactions overall is relatively small, for some compounds such as acetonitrile and dimethylsulfoxide these contributions reach nearly 20 %. In these cases, inductive interactions can be about 1/3 as important as dispersive interactions.

The measurement of $L^{16}$ can be very difficult, especially for very large or very polar compounds. Column preparation is difficult, and even at 25 °C, a significant amount of hexadecane will bleed from the column. Additionally, gas-solid interfacial adsorption may

contribute to the retention of a compound. Also, the low temperature makes it impossible to measure $L^{16}$ for very large compounds. In this work, we have shown that it is possible to determine $\log L^{16}$ within 0.05 by measuring retention on commercially available, polymeric stationary phases at high temperatures. Although retention on several stationary phases is required (7 in this study), these measurements are generally rapid and precise.

## Analysis of Gas Chromatographic Retention Using $V_X$, $P_X$, $D_X$, and $A_X$

We can also use the new descriptors to correlate the GC retention data used in this study. The coefficient of determination values were all greater than 0.996 for the 29 conditions studied (7 different stationary phases at 45 to 150 °C). The good fits to GC retention data are expected since some of this data was used to define the descriptors, and thus provide no evidence that the new parameters are superior to other LSER parameters. However, the signs and magnitudes of these coefficients make chemical sense, indicating that the parameters do reflect chemical interaction ability. Additionally, the magnitudes of the coefficients, which reflect the strength of the interactions governing retention, decrease with increasing temperature as expected. The *d* coefficient, which is a measure of the stationary phase's dipolarity and polarizability, increases with the dipolarity of the stationary phase. Furthermore, the *d* coefficients from this study scale with the *s* coefficients found using equation 1 [11]. Since equation 1 is an accepted model for GC retention, the good agreement of the *s* coefficient with our *d* coefficient provides considerable support for the validity of our model and approach.

14

<u>Test of $D_X$ as a Solute Dipolarity Parameter</u>

Using the current approach, we have proposed that it is possible to define a solute

dipolarity parameter ($D_X$) free of any polarizability contribution.  To test this dipolarity

parameter we compare it to various other dipolarity and polarizability parameters, including

$*_2^C$,  $_2^H$, $R_2$, and the solute dipole moment ($\mu$).  Since $D_X$ was calculated from some of the

same GC retention data used to measure  $*_2^C$, we first test to see if there is a significant

difference between $D_X$ and  $*_2^C$.  $D_X$ is related to  $*_2^C$ as follows:

$$D_X = \left(0.089 \pm 0.024\right) + \left(0.734 \pm 0.052\right) \ *_2^C \tag{28}$$

$$R = 0.892 \qquad \overline{sd} = 0.12 \qquad N = 53.$$

$D_X$ correlates slightly better with  $_2^H$, as follows:

$$D_X = \left(0.609 \pm 0.019\right) \ _2^H \tag{29}$$

$$R = 0.927 \qquad \overline{sd} = 0.096 \qquad N = 53.$$

Figure 5 shows the comparison between $D_X$ and  $_2^H$.  Note that all of the aromatic

compounds fall below the regression line. Cyclohexane and carbon  tetrachloride, which

have high  polarizability-to-volume  ratios,  also  fall  below  the  regression  line.   The

fluorinated compounds, which have small polarizability-to-volume  ratios,  fall  above  the

regression line.  Clearly the polarizability contribution of $D_X$ is different from that of  $_2^H$.

If $D_X$ is a true dipolarity parameter, it should correlate fairly well with the square of

the dipole moment, $\mu^2$:

$$D_X = \left(0.131 \pm 0.020\right) + \left(0.0465 \pm 0.0029\right) \mu^2 \tag{30}$$

$$R = 0.914 \qquad \overline{sd} = 0.10 \qquad N = 53.$$

Although the quality of this fit is fairly poor, both $*_2^C$ and $_2^H$ correlate much more poorly with $\mu^2$:

$$*_2^C = \left(0.127 \pm 0.037\right) + \left(0.0480 \pm 0.0055\right) \mu^2 \tag{31}$$

$$R = 0.776 \qquad \overline{sd} = 0.20 \qquad N = 53,$$

$$_2^H = \left(0.264 \pm 0.040\right) + \left(0.0651 \pm 0.0060\right) \mu^2 \tag{32}$$

$$R = 0.836 \qquad \overline{sd} = 0.22 \qquad N = 53.$$

According to these regressions, $D_X$ appears to be a better pure dipolarity descriptor than $*_2^C$ and $_2^H$.

Both $*_2^C$ and $_2^H$ are considered to include a significant polarizability contribution. The correlations in equation 31 and 32 should therefore improve with the addition of Abraham's polarizability parameter, $R_2$. We find a significant improvement in the quality of the fits, as shown in the following equations:

$$*_2^C = \left(0.0414 \pm 0.0029\right) \mu^2 + \left(0.524 \pm 0.045\right) R_2 \tag{33}$$

$$R = 0.930 \qquad \overline{sd} = 0.11 \qquad N = 53,$$

$$_2^H = \left(0.116 \pm 0.031\right) + \left(0.0592 \pm 0.0039\right) \mu^2 + \left(0.565 \pm 0.065\right) R_2 \tag{34}$$

$$R = 0.938 \qquad \overline{sd} = 0.14 \qquad N = 53.$$

However, it can be seen by comparing equation 31 with equation 35 below, that the quality of the fit for $D_X$ does not increase significantly when $R_2$ is included in the regression. This indicates that the contribution of polarizability to $D_X$ relative to the dipolarity contribution is very small.

$$D_X = \left(0.100 \pm 0.022\right) + \left(0.0452 \pm 0.0028\right) \mu^2 + \left(0.116 \pm 0.047\right) R_2 \qquad (35)$$

$$R = 0.924 \qquad \overline{sd} = 0.10 \qquad N = 53.$$

In equation 33 and 34, the ratio of the polarizability coefficient to the dipolarity coefficient is 12.6 for $*_2^C$ and 9.6 for $_2^H$. (This does not mean that polarizability is 12.6 or 9.6 times more important than dipolarity for $*_2^C$ and $_2^H$. We calculate this ratio only for comparisons between the three scales.) The same ratio for $D_X$ is only 2.6. From these results, we conclude that there is little or no polarizability contribution to $D_X$. If there is a small contribution, it is considerably smaller than that for $*_2^C$ or $_2^H$.

Generality of the New Parameters

The new parameters, $V_X$, $P_X$, $D_X$, and $A_X$ are determined from GC retention data, and they give excellent fits to such data, as described above. However, for these parameters to be considered universal, they must be able to correlate additional types of solution processes. An important set of data that has previously been studied by LSERs is the set of octanol-water partition coefficients ($K_{OW}$) [3]. These values are important because they are used for the development of quantitative structure/activity relationships in medicinal

chemistry, and they can be used to predict properties such as solubility and retention in RPLC. Correlations of log $K_{OW}$ require a solute basicity parameter [3]. We use Abraham's $\beta_2^H$ together with $V_X$, $P_X$, $D_X$, and $A_X$ to obtain the following correlation:

$$logK_{OW} = (0.0475 \pm 0.0046) V_X - (0.082 \pm 0.037) P_X - (1.90 \pm 0.13) D_X - (0.19 \pm 0.11) A_X - (2.97 \pm 0.12) \beta_2^H \quad (36)$$

$$R = 0.997 \qquad \overline{sd} = 0.16 \qquad N = 46.$$

The correlation of log $K_{OW}$ is not a severe test of the generality of the parameters, since $K_{OW}$ is only weakly dependent on dipolarity and hydrogen bond acidity. However, the fairly good quality of fit for this regression demonstrates that these parameters can be used to correlate water-solvent partitions in addition to GC retention. For LSERs involving the conventional sets of parameters (either Abraham or Carr parameters), two different equations (differing in the dispersion/cavity formation term, log $L^{16}$ or $V_X$) must be used, along with a floating intercept. Comparing equation 36 with Abraham's LSER for log $K_{OW}$,

$$logK_{OW} = (0.154 \pm 0.094) + (3.826 \pm 0.076) V_M + (0.588 \pm 0.091) R_2$$
$$- (1.150 \pm 0.079) \pi_2^H - (0.140 \pm 0.097) \alpha_2^H - (3.43 \pm 0.10) \beta_2^H \quad (37)$$

$$R = 0.998 \qquad \overline{sd} = 0.13 \qquad N = 46,$$

we find that equation 36 does not fit the data as well as equation 37. We expect this to be the case for all processes involving water as a solvent, because Abraham calculated his parameters from a much wider range of data. For predictive purposes, or goodness of fit, it is unlikely that we will be able to do better than Abraham. However, in terms of interpreting the data, we believe that equation 36 is more useful than equation 37.

<u>Interpretation of LSER Coefficients</u>

We can use the same set of parameters to correlate gas-to-water partition coefficients ($L^W$) as follows:

$$logL^W = \left(-0.1042\pm 0.0065\right) V_X + \left(0.739\pm 0.052\right) P_X + \left(4.24\pm 0.18\right) D_X$$
$$+\left(3.46\pm 0.17\right) A_X + \left(5.08\pm 0.18\right) \quad {}_2^H \tag{38}$$

$$R = 0.994 \qquad \overline{sd} = 0.24 \qquad N = 47.$$

The signs and magnitudes of the coefficients match what we would expect from previous studies of gas-to-water partitioning [4]. Again, however, the quality of this fit is not as good as it is using Abraham's parameters, [4]

$$log\ L^W = \left(-0.72\pm 0.11\right) - \left(1.10\pm 0.10\right) V_M + \left(0.694\pm 0.091\right) R_2$$
$$+\left(2.346\pm 0.81\right) {}_2^H + \left(3.51\pm 0.11\right) {}_2^H + \left(4.90\pm 0.10\right) {}_2^H \tag{39}$$

$$R = 0.998 \qquad \overline{sd} = 0.14 \qquad N = 47.$$

We note, though, that the interpretation of equation 38 is much less complex than that of equation 40. First, equation 39 includes an intercept, which is difficult to explain from the perspective of the cavity model of solvation. Without this intercept, the quality of the fit decreases significantly. Second, the coefficient of volume in equation 39 includes both cavity formation and dispersion contributions, whereas the same coefficient in equation 38 is purely a cavity term. Finally, the coefficient of the $R_2$ parameter in equation 39 is difficult to explain in terms of either dispersion or induction interactions.

Although we believe we have succeeded in defining a solute dipolarity term with little or no dependence on polarizability, the interpretation of the *coefficients* of $P_X$ and $D_X$

is still complex.  From earlier work, we know that the coefficient of $P_X$ represents  the

solvent's polarizability (reflecting the contribution of dispersion interactions) and dipolarity

(induction interactions with a dipolar solvent), and that the coefficient of $D_X$ represents the

solvent's dipolarity (dipole-dipole interactions) and polarizability  (induction  interactions

with a dipolar solute).

This complexity is evident if we consider the LSER for the  water-to-hexadecane

partition coefficient ($P_{W-C16}$), which we can calculate from equations 27 and 38:

$$log\ P_{W-C16} = 0.0586\ V_X - 0.119\ P_X - 2.72\ D_X - 3.46\ A_X - 5.08\ _2^H \qquad (40)$$

According  to  the  negative  $p$  coefficient  in  equation  40,  the  combination  of  dispersion

interactions and induction interactions with a dipolar solvent are more favorable for water

than they are for hexadecane.  It is doubtful that water, with its particularly low index of

refraction,  can  interact  dispersively  with  solute  molecules  better  than  hexadecane  can.

Therefore,  the  small  negative  $p$  coefficient  in  equation  40  must  be  due  to  induction

interactions between the solvent dipole and the solute induced dipole.

Correlations of RPLC Retention

As a final test of our parameters, we were able to correlate retention in reversed-

phase liquid chromatography (RPLC) for several binary mobile phases containing water,

acetonitrile, methanol, and tetrahydrofuran in different mobile phase  compositions.    An

example of the correlations obtained is given in equation 41:

$$log\,k_{RPLC,\,50/50\,MeOH} = \left(0.74 \pm 0.15\right) + \left(0.0376 \pm 0.0072\right) V_X - \left(0.116 \pm 0.047\right) P_X$$
$$- \left(1.18 \pm 0.12\right) D_X - \left(0.375 \pm 0.083\right) A_X - \left(1.55 \pm 0.12\right) \beta_2^H \quad (41)$$

$$R = 0.996 \qquad \overline{sd} = 0.082 \qquad N = 21,$$

This equation, combined with the LSERs for $L^{16}$, $K_{OW}$, and $L^W$ presented above indicate that the parameters derived in this work can be used to correlate and explain a variety of gas-liquid and liquid-liquid processes, demonstrating the generality of the new parameters.

## Conclusions

Using the theoretical basis for LSERs we derived a set of solute descriptors including volume, polarizability, dipolarity, and hydrogen bond acidity. These parameters are based only on GC retention data, based on only a few assumptions. In particular, a solute dipolarity term was developed which has little or no dependence on solute polarizability. Although these parameters generally produce poorer fits than the Abraham parameters for systems involving water, the interpretation of the coefficients is much more straightforward. Several correlations were presented and chemically interpreted.

## Acknowledgments

## References

[1]   M. J. Kamlet and R. W. Taft, *Acta Chem. Scand. B 39*, (1985), 611.

[2]   M. H. Abraham and R. Fuchs, *J. Chem. Soc. Perkin Trans. 2*, (1988), 523.

[3]   M. H. Abraham, H. S. Chadha, G. S. Whiting, and R. C. Mitchell, *J. Pharmaceutical Sciences*, 83 (1994) 1085.

[4]   M. H. Abraham, G. S. Whiting, R. Fuchs, and E. J. Chambers, *J. Chem. Soc. Perkin Trans. 2*, (1990) 291.

[5]   M. J. Kamlet and R. W. Taft *J. Chem. Soc. Perkin Trans. 2*, (1979) 337.

[6]   M. H. Abraham, *Chem. Soc. Rev.*, 22 (1993) 73.

[7]   M. H. Abraham and M. Roses, *J. Phys. Org. Chem.*, 7 (1994) 672.

[8]   M. H. Abraham, H. S. Chadha, and J. P. Dixon, *J. Phys. Org. Chem.*, 7 (1994) 712.

[9]   M. H. Abraham and A. Nasehzadeh, *J. Chem. Soc. Faraday Trans. 1*, 77 (1981) 321.

[10]  J. C. Dearden, S. J. A. Bradburne, and M. H. Abraham, *QSAR:  Rational Approaches to the Design of Bioactive Compounds*, (1991), 143.

[11]  J. D. Weckwerth, *Solvatochromic and Thermodynamic Studies of Retention in Supercritical Fluid Chromatography*, Ph. D. Thesis, University of Minnesota (1996).

[12]  J. Li, Y. Zhang, A. J. Dallas, and P. W. Carr, *J. Chromatogr.*, 550 (1991) 101.

[13]  Abraham, M. H., Whiting, G. S., Doherty, R. M., Shuely, W. J., *J. Chromatogr.*, 587 (1991) 213.

[14]  E. F. Meyer and R. E. Wagner, *J. Phys. Chem.*, 70, (1966) 3162.

[15]  J. E. Brady and P. W. Carr, *J. Phys. Chem.* 86 (1982) 3053.

[16]  J. E. Brady and P. W. Carr, *J. Phys. Chem.* 89 (1985) 1813.

[17]  J. Li, A. J. Dallas, and P. W. Carr, *J. Chromatogr.*, 517 (1990) 103.

[18]  P. J. Rousseeuw and A. M.Leroy, "Robust Regression and Outlier Detection", John
      Wiley & Sons, New York, New York, 1987.

[19]  MATLAB for Windows, Version 4.2c.1, Copyright 1984-1994, The Mathworks, Inc.

[20]  M. H. Abraham and J. C. McGowan, *Chromatographia*, 23 (1987) 243.

[21]  K. J. Miller and J. A. Savchik, *J. Am. Chem. Soc.*, (1979) 7206.

[22]  A. J. Dallas and P. W. Carr, *J. Phys. Chem.*, 98 (1994) 4927.

Table 1.  Solute Volume ($V_X$), Polarizability ($P_X$), Dipolarity ($D_X$), and Acidity ($A_X$)

Parameters.

| Compound | $V_X$ | $P_X$ | $D_X$ | $A_X$ |
|---|---|---|---|---|
| cyclohexane | 96.42 | 11.90 | 0.00 | 0.00 |
| 1-hexene | 95.39 | 11.18 | 0.02 | 0.00 |
| pentane | 87.87 | 9.98 | 0.00 | 0.00 |
| hexane | 101.96 | 11.83 | 0.00 | 0.00 |
| octane | 130.14 | 15.52 | 0.00 | 0.00 |
| decane | 158.32 | 19.22 | 0.00 | 0.00 |
| undecane | 172.41 | 21.06 | 0.00 | 0.00 |
| tetradecane | 214.68 | 26.60 | 0.00 | 0.00 |
| pentadecane | 228.77 | 28.45 | 0.00 | 0.00 |
| ethyl acetate | 75.00 | 8.38 | 0.39 | 0.00 |
| propyl acetate | 90.87 | 10.38 | 0.40 | 0.00 |
| diethyl ether | 72.89 | 8.50 | 0.08 | 0.00 |
| dipropyl ether | 100.94 | 12.16 | 0.06 | 0.00 |
| dibutyl ether | 126.24 | 15.59 | 0.06 | 0.00 |
| acetonitrile | 44.74 | 4.27 | 0.66 | 0.00 |
| propionitrile | 57.53 | 5.88 | 0.66 | 0.00 |
| acetone | 58.42 | 5.65 | 0.56 | 0.00 |
| butanone | 75.30 | 7.88 | 0.54 | 0.00 |
| 2-pentanone | 85.57 | 9.34 | 0.53 | 0.00 |
| *N,N*-dimethylformamide | 76.99 | 8.28 | 0.85 | 0.00 |
| *N,N*-dimethylacetamide | 89.42 | 9.91 | 0.85 | 0.00 |
| dimethylsulfoxide | 74.57 | 8.00 | 1.00 | 0.00 |
| propionaldehyde | 55.51 | 5.76 | 0.45 | 0.00 |
| tetrahydrofuran | 72.71 | 8.59 | 0.31 | 0.00 |
| triethylamine | 103.67 | 12.03 | 0.16 | 0.00 |
| nitromethane | 49.77 | 5.10 | 0.68 | 0.00 |
| nitroethane | 63.39 | 6.82 | 0.67 | 0.00 |
| 1-nitropropane | 77.56 | 8.65 | 0.66 | 0.00 |
| methanol | 29.50 | 3.14 | 0.25 | 0.31 |
| ethanol | 44.27 | 4.82 | 0.29 | 0.24 |
| 1-propanol | 63.13 | 7.10 | 0.31 | 0.24 |

| | | | | |
|---|---|---|---|---|
| 2-propanol | 58.40 | 6.37 | 0.29 | 0.22 |
| 2-methyl-2-propanol | 64.35 | 7.08 | 0.30 | 0.14 |
| 2,2,2-trifluoroethanol | 53.60 | 4.66 | 0.48 | 0.59 |
| 1,1,1,3,3,3-hexafluoro-2-propanol | 80.25 | 6.88 | 0.52 | 1.00 |
| acetic acid | 47.48 | 5.27 | 0.45 | 0.46 |
| aniline | 89.79 | 11.77 | 0.46 | 0.22 |
| *N*-methylaniline | 105.30 | 13.77 | 0.43 | 0.16 |
| phenol | 90.99 | 11.74 | 0.39 | 0.67 |
| benzylalcohol | 95.96 | 12.62 | 0.43 | 0.35 |
| m-cresol | 104.54 | 13.46 | 0.42 | 0.63 |
| ethylamine | 57.17 | 5.99 | 0.31 | 0.00 |
| propylamine | 69.31 | 7.56 | 0.34 | 0.00 |
| butylamine | 83.60 | 9.33 | 0.38 | 0.00 |
| benzene | 78.21 | 9.80 | 0.18 | 0.00 |
| toluene | 93.16 | 11.69 | 0.20 | 0.00 |
| ethylbenzene | 104.48 | 13.30 | 0.18 | 0.00 |
| propylbenzene | 117.71 | 15.00 | 0.18 | 0.00 |
| *p*-xylene | 108.04 | 13.64 | 0.19 | 0.00 |
| benzaldehyde | 95.91 | 12.12 | 0.53 | 0.00 |
| benzonitrile | 100.08 | 12.21 | 0.65 | 0.00 |
| *N,N*-dimethylaniline | 119.88 | 15.47 | 0.38 | 0.00 |
| carbon tetrachloride | 82.96 | 10.58 | 0.06 | 0.00 |

## Figure Captions

Figure 1.  Comparison of $A_X$ with ${}_2^H$ and ${}_2^C$.  □) $A_X$ vs. ${}_2^H$; ○) $A_X$ vs. ${}_2^C$.  The line is the linear regression of $A_X$ vs. ${}_2^C$.

Figure 2.  Comparison of $V_X$ and $V_M'$.  The line is the linear regression of $V_X$ vs. $V_M'$.

Figure 3.  Comparison of $P_X$ and $P_{M-S}$.  The line is the linear regression of $P_X$ vs. $P_{M-S}$.

Figure 4.  Calculated vs. Experimental $\log L^{16}$.  Calculated values were determined using equation 27 and the parameters in Table 1.

Figure 5.  Comparison of $D_X$ and ${}_2^H$.  ■) alkanes;  ) cyclohexane; ▲) aromatic compounds; ●) fluorinated compounds; ▼) carbon tetrachloride; and ○) all other solutes.  The line is the linear regression of $D_X$ vs. ${}_2^H$.

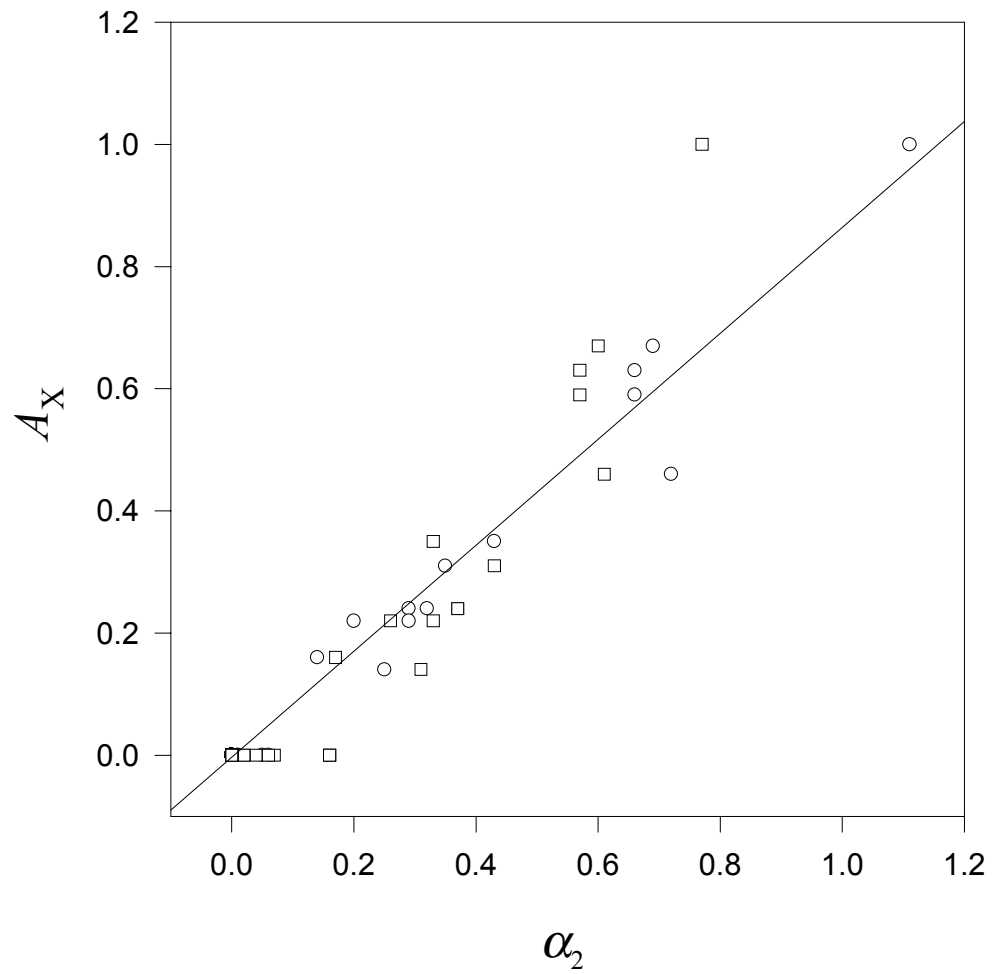**Figure 1.  Comparison of $A_X$ with $\Sigma\alpha_2^H$ and $\alpha_2^C$.**
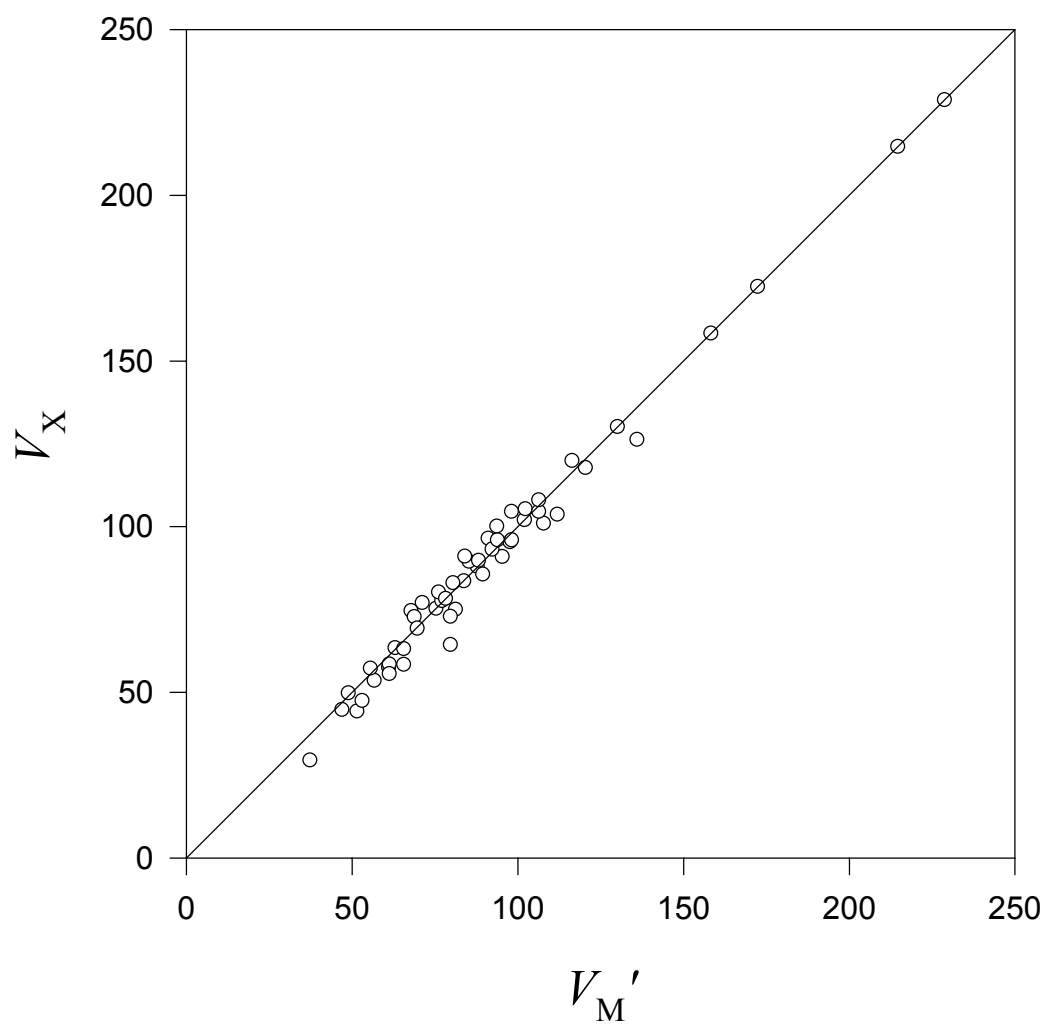
**Figure 2. Comparison of $V_X$ and $V_m'$.**

**Figure 3.  Comparison of $P_X$ and $P_{M-S}$.**
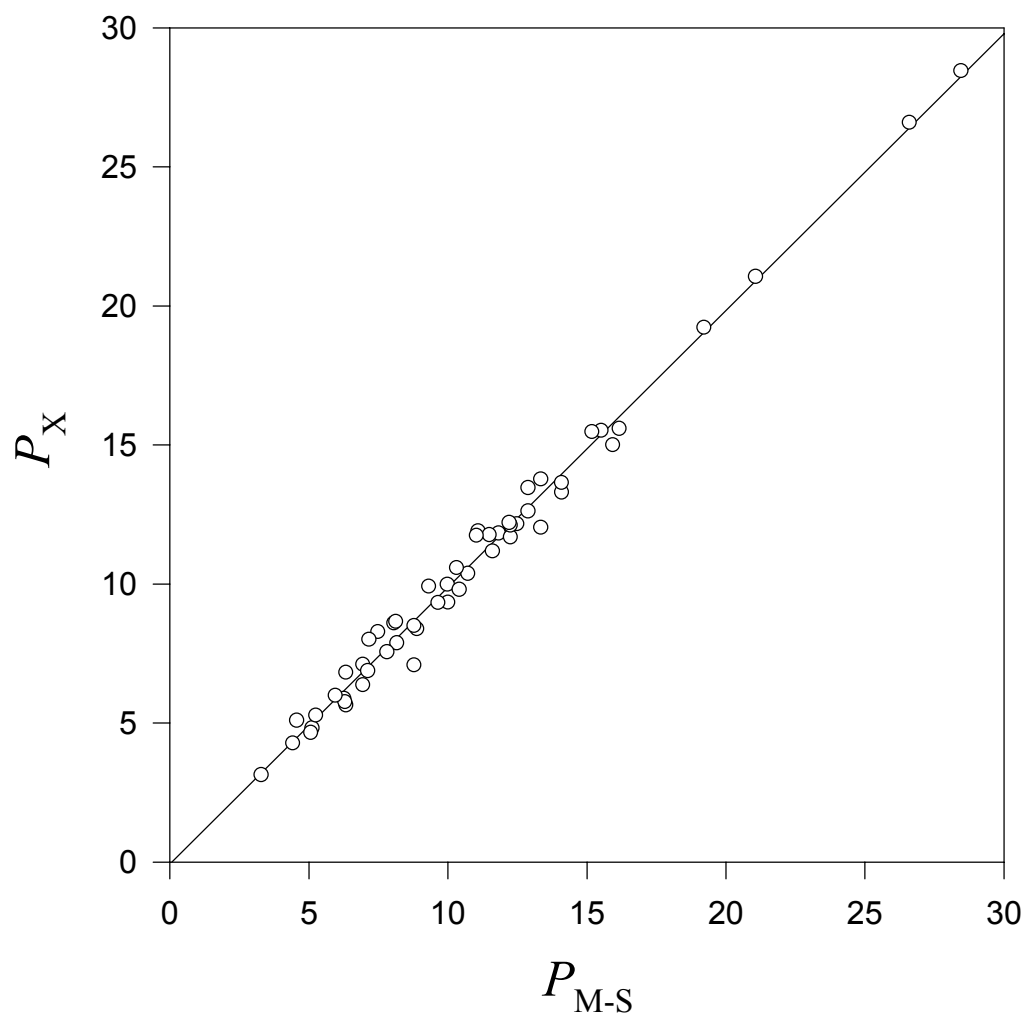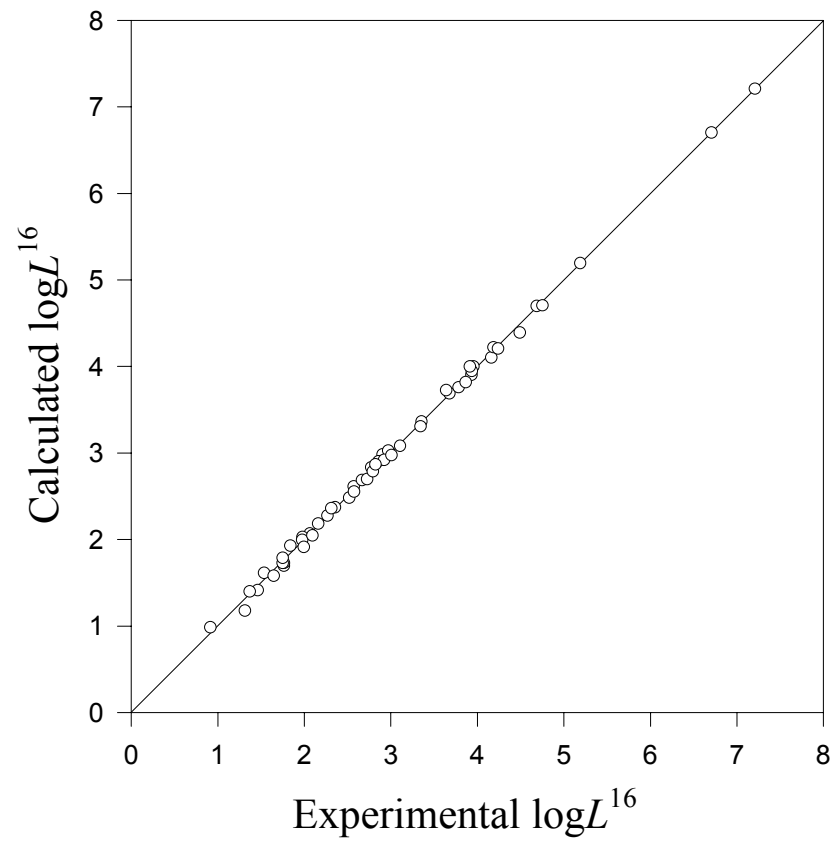
**Figure 4.  Calculated vs. Experimental log$L^{16}$.**



Figure 4.  Calculated vs. Experimental log$L^{16}$.

**Figure 5.  Comparison of $D_X$ and $\pi_2{}^H$.**